On mixing times for stratified walks on the d-cube

Nancy Lopes Garcia

Universidade Estadual de Campinas

Campinas, Brasil

and

José Luis Palacios

Universidad Simón Bolívar

Caracas, Venezuela.

February 2, 2008

Abstract

Using the electric and coupling approaches, we derive a series of results concerning the mixing times for the stratified random walk on the d-cube, inspired in the results of Chung and Graham (1997) Stratified random walks on the n-cube. *Random Structures and Algorithms*, **11**,199-222.

Key Words: effective resistance, coupling, birth and death chains

1991 Mathematics Subject Classification. Primary: 60J15; secondary: 60C05.

1 Introduction.

The stratified random walk (SRW) on the d-cube Q_d is the Markov chain whose state space is the set of vertices of the d-cube and whose transition probabilities are defined thus:

Given a set of non-zero probabilities $p = (p_0, p_1, \ldots, p_{d-1})$, from any vertex with k 1's, the process moves either to any neighboring vertex with $k+1$ 1's with probability $\frac{p_k}{d}$; or to any neighboring vertex with $k-1$ 1's with probability $\frac{p_{k-1}}{d}$; or to itself with the remaining probability. The simple random walk on the d-cube corresponds to the choice $p_k = 1$ for all k.

Vaguely speaking, the mixing time of a Markov chain is the time it takes the chain to have its distribution close to the stationary distribution under some measure of closeness. Chung and Graham studied the SRW on the d-cube in [5], mainly with algebraic methods, and found bounds for the mixing times under total variation and relative pointwise distances. Here we use non-algebraic methods, the electric and coupling approaches, in order to study the same SRW and get exact results for maximal commute times and bounds for cover times and mixing times under total variation distance. We take advantage of the fact that there seems to be some inequality or another linking hitting times, commute times, cover times and any definition of mixing time with any other under any measure of closeness (see Aldous and Fill [2] and Lovász and Winkler [8]).

2 The electric approach

On a connected undirected graph $G = (V, E)$ such that the edge between vertices i and j is given a resistance r_{ij} (or equivalently, a conductance $C_{ij} = 1/r_{ij}$), we can define the random walk on G as the Markov chain $\mathbf{X} = \{\mathbf{X}(n)\}_{n \geq 0}$ that from its current vertex v jumps to the neighboring vertex w with probability $p_{vw} = C_{vw}/C(v)$, where $C(v) = \sum_{w:w\sim v} C_{vw}$, and $w \sim v$ means that w is a neighbor of v. There may be a conductance C_{zz} from a vertex z to itself, giving rise to a transition probability from z to itself. Some notation: $\mathbb{E}_a T_b$ and $\mathbb{E}_a C$ denote the expected value, starting from the vertex a, of respectively, the hitting time T_b of the vertex

b and the cover time C, i. e., the number of jumps needed to visit all the states in V; R_{ab} is the effective resistance, as computed by means of Ohm's law, between vertices a and b.

A Markov chain is reversible if $\pi_i \mathbb{P}(i,j) = \pi_j \mathbb{P}(j,i)$ for all i, j, where $\{\pi_\cdot\}$ is the stationary distribution and $\mathbb{P}(\cdot,\cdot)$ are the transition probabilities. Such a reversible Markov chain can be described as a random walk on a graph if we define conductances thus:

$$C_{ij} = \pi_i \mathbb{P}(i,j). \tag{2.1}$$

We will be interested in finding a closed form expression for the commute time $\mathbb{E}_0 T_d + \mathbb{E}_d T_0$ between the origin, denoted by 0, and its opposite vertex, denoted by d.

Notice first that the transition matrix for $\mathbf{X} = \{\mathbf{X}(n), n \geq 0\}$, the SRW on the d-cube, is doubly stochastic and therefore its stationary distribution is uniform. If we now collapse all vertices in the cube with the same number of 1's into a single vertex, and we look at the SRW on this collapsed graph, we obtain a new *reversible* Markov chain $\mathbf{S} = \{\mathbf{S}(n), n \geq 0\}$, a birth-and-death chain in fact, on the state space $\{0, 1, \ldots d\}$, with transition probabilities

$$\begin{aligned} \mathbb{P}(k,k+1) &= \frac{d-k}{d} p_k, & (2.2)\\ \mathbb{P}(k,k-1) &= \frac{k}{d} p_{k-1}, & (2.3)\\ \mathbb{P}(k,k) &= 1 - \mathbb{P}(k,k+1) - \mathbb{P}(k,k-1). & (2.4) \end{aligned}$$

It is plain to see that the stationary distribution of this new chain is the Binomial with parameters d and $\frac{1}{2}$. It is also clear that the commute time between vertices 0 and d is the same for both $\mathbf{X}$ and $\mathbf{S}$. For the latter we use the electric machinery described above, namely, we think of a linear electric circuit from 0 to d with conductances given by (2.1) for $0 \leq i \leq d$, $j = i-1, i, i+1$, and where $\pi_i = \binom{d}{i} \frac{1}{2^d}$.

It is well known (at least since Chandra et al. proved it in [4]) that

$$\mathbb{E}_a T_b + \mathbb{E}_a T_b = R_{ab} \sum_z C(z), \tag{2.5}$$

where R_{ab} is the effective resistance between vertices a and b.

If this formula is applied to a reversible chain whose conductances are given as in (2.1), then it is clear that

$$C(z) = \pi_z$$

and therefore the summation in (2.5) equals 1. We get then this compact formula for the commute time:

$$\mathbb{E}_a T_b + \mathbb{E}_a T_b = R_{ab}, \tag{2.6}$$

where the effective resistance is computed with the individual resistors having resistances

$$r_{ij} = \frac{1}{C_{ij}} = \frac{1}{\pi_i \mathbb{P}(i,j)}.$$

In our particular case of the collapsed chain, because it is a linear circuit, the effective resistance R_{0d} equals the sum of all the individual resistances $r_{i,i+1}$, so that (2.6) yields

$$\mathbb{E}_0 T_d + \mathbb{E}_d T_0 = R_{0d} = 2^d \sum_{k=0}^{d-1} \frac{1}{p_k \binom{d-1}{k}}. \tag{2.7}$$

Because of the particular nature of the chain under consideration, it is clear that $\mathbb{E}_0 T_d + \mathbb{E}_d T_0$ equals the maximal commute time (τ^* in the terminology of Aldous [2]) between any two vertices.

(i) For simple random walk, formula (2.7) is simplified by taking all $p_k = 1$. This particular formula was obtained in [8] with a more direct argument, and it was argued there that

$$\sum_{k=0}^{d-1} \frac{1}{\binom{d-1}{k}} = 2 + o(1).$$

An application of Matthews' result (see [9]), linking maximum and minimum expected hitting times with expected cover times, yields immediately that the expected cover time is $\mathbb{E}_v C = \Theta(|V| \log |V|)$, which is the asymptotic value of the lower bound for cover times of walks on a graph $G = (V, E)$ (see [6]). Thus we could say this SRW is a "rapidly covered" walk.

(ii) The so-called Aldous cube (see [5]) corresponds to the choice $p_k = \frac{k}{d-1}$. This walk takes place in the "punctured cube" that excludes the origin. Formula (2.7) thus, must exclude $k = 0$ in this case, for which we still get a closed-form expression for the commute time between vertex

d, all of whose coordinates are 1, and vertex s which consists of the collapse of all vertices with a single 1:

$$\mathbb{E}_s T_d + \mathbb{E}_d T_s = 2^d \sum_{k=1}^{d-1} \frac{1}{\binom{d-2}{k-1}}. \tag{2.8}$$

The same argument used in (i) tells us that the summation in (2.8) equals $2 + o(1)$ and, once again, Matthews' result tells us that the walk on the Aldous cube has a cover time of order $|V| \log |V|$.

(iii) The choice $p_k = \frac{1}{\binom{d-1}{k}}$ would be in the terminology of [5] a "slow walk": the commute time is seen to be exactly equal to $|V| \log_2 |V|$ and thus the expected cover time is $O(|V| \log^2 |V|)$. In general, the SRW will be rapidly covered if and only if

$$\sum_{k=0}^{d-1} \frac{1}{p_k \binom{d-1}{k}} = c + o(1),$$

for some constant c.

Remark. A formula as compact as (2.7) could be easily obtained through the commute time formula (2.6). It does not seem that it could be obtained that easily, by just adding the individual hitting times $\mathbb{E}_i T_{i+1}$. (A procedure that is done, for instance, in [5], [10], [11], and in the next section).

3 The coupling approach

In order to asess the rate of convergence of the SRW on the cube Q_d to the uniform stationary distribution π, we will bound the mixing time τ defined as

$$\tau = \min\{t : d(t') \leq \frac{1}{2e}, \text{ for all } t' > t\},$$

where

$$d(t) = \max_{\mathbf{x} \in Q_d} \|P_{\mathbf{x}}(\mathbf{X}(t) = \cdot) - \pi(\cdot)\|,$$

and $\|\theta_1 - \theta_2\|$ is the variation distance between probability distributions θ_1 and θ_2, one of whose alternative definitions is (see Aldous and Fill [2]), chapter 2):

$$\|\theta_1 - \theta_2\| = \min \mathbb{P}(V_1 \neq V_2),$$

where the minimum is taken over random pairs (V_1, V_2) such that V_m has distribution $\theta_m, m = 1, 2$.

The bound for the mixing time is achieved using a coupling argument that goes as follows: let $\{\mathbf{X}(t), t \geq 0\}$ and $\{\mathbf{Y}(t), t \geq 0\}$ be two versions of the SRW on Q_d such that $\mathbf{X}(0) = \mathbf{x}$ and $\mathbf{Y}(0) \sim \pi$. Then

$$\|\mathbb{P}_{\mathbf{x}}(\mathbf{X}(t) = \cdot) - \pi(\cdot)\| \leq \mathbb{P}(\mathbf{X}(t) \neq \mathbf{Y}(t)). \tag{3.1}$$

A coupling between the processes $\mathbf{X}$ and $\mathbf{Y}$ is a bivariate process such that its marginals have the distributions of the original processes and such that once the bivariate process enters the diagonal, it stays there forever. If we denote by

$$T_{\mathbf{x}} = \inf\{t; \mathbf{X}(t) = \mathbf{Y}(t)\}$$

the coupling time, i. e., the hitting time of the diagonal, then (3.1) translates as

$$\|\mathbb{P}_{\mathbf{x}}(\mathbf{X}(t) = \cdot) - \pi(\cdot)\| \leq \mathbb{P}(T_{\mathbf{x}} > t), \tag{3.2}$$

and therefore,

$$d(t) \leq \max_{\mathbf{x} \in Q_d} \mathbb{P}(T_{\mathbf{x}} > t). \tag{3.3}$$

If we can find a coupling such that $\mathbb{E}T_{\mathbf{x}} = O(f(d))$, for all $x \in Q_d$ and for a certain function f of the dimension d, then we will also have $\tau = O(f(d))$. Indeed, if we take $t = 2ef(d)$, then (3.3) and Markov's inequality imply that $d(t) \leq 1/2e$ and the definition of τ implies $\tau = O(f(d))$.

We will split $T_{\mathbf{x}}$ as $T_{\mathbf{x}} = T^1_{\mathbf{x}} + T^2_{\mathbf{x}}$, where $T^1_{\mathbf{x}}$ is a coupling time for the birth-and-death process $\mathbf{S}$, and $T^2_{\mathbf{x}}$ is another coupling time for the whole process $\mathbf{X}$, once the bivariate $\mathbf{S}$ process enters the diagonal, and we will bound the values of $\mathbb{E}T^1_{\mathbf{x}}$ and $\mathbb{E}T^2_{\mathbf{x}}$.

More formally, for any $\mathbf{x}, \mathbf{y} \in Q_d$ define

$$s(\mathbf{x}) = \sum_{i=1}^{d} x_i \tag{3.4}$$

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{d} |x_i - y_i|. \tag{3.5}$$

Define also for the birth-and-death process $\mathbf{S}(t) = s(\mathbf{X}(t))$ its own mixing time:

$$\tau^{(S)} = \inf\{t; d_S(t) \leq \frac{1}{2e}\},$$

where $d_S(t) = \max_i \|\mathbb{P}_i(\mathbf{S}(t) = \cdot) - \pi_S(\cdot)\|$, and π_S is the stationary distribution of $\mathbf{S}$. Notice that $s(\mathbf{Y}(0)) \sim \pi_S$ since $\mathbf{Y}(0) \sim \pi$.

Now we will prove that $\tau^{(S)} = O(f_S(d))$, for a certain function f_S of the expected hitting times of the "central states", and that this bound implies an analogous bound for $\mathbb{E}T^1_{\mathbf{x}}$. Indeed, as shown by Aldous [1], we can bound $\tau^{(S)}$ by a more convenient stopping time

$$\tau^{(S)} \leq K_2 \tau_1^{(3)} \tag{3.6}$$

where $\tau_1^{(3)} = \min_\mu \max_i \min_{U_i} \mathbb{E}_i U_i$ and the innermost minimum is taken over stopping times U_i such that $\mathbb{P}_i(S(U_i) \in \cdot) = \mu(\cdot)$. In particular,

$$\begin{aligned}
\tau_1^{(3)} &\leq \min_b \max_i \min_{U_i^b} \mathbb{E}_i\, U_i^b && (3.7)\\
&\leq \min_b \max_i \mathbb{E}_i T_b && (3.8)\\
&= \max(\mathbb{E}_0 T_{d/2}, \mathbb{E}_d T_{d/2}) && (3.9)
\end{aligned}$$

where the innermost minimum in (3.7) is taken over stopping times U_i^b such that $\mathbb{P}_i(S(U_i) = b) = 1$. Expression (3.9) follows from (3.8) since we are dealing with birth and death chains. Therefore, combining (3.6) and (3.9) we have

$$\tau^{(S)} \leq K_2 \, \max(\mathbb{E}_0 T_{d/2}, \mathbb{E}_d T_{d/2}) := f_S(d). \tag{3.10}$$

In general, a coupling implies an inequality like (3.2). However, the inequality becomes an equality for a certain maximal (non-Markovian) coupling, described by Griffeath [7]. Let $T^1_{\mathbf{x}}$ be the coupling time for the maximal coupling between $s(\mathbf{X}(t))$ and $s(\mathbf{Y}(t))$ such that

$$\|P_{\mathbf{x}}(S(\mathbf{X}(t)) = \cdot) - \pi_S(\cdot)\| = \mathbb{P}(T^1_{\mathbf{x}} > t).$$

Then

$$d_S(t) = \max_{\mathbf{x} \in Q_d} \|P_{\mathbf{x}}(S(\mathbf{X}(t)) = \cdot) - \pi_S(\cdot)\| = \max_{\mathbf{x} \in Q_d} \mathbb{P}(T^1_{\mathbf{x}} > t).$$

By the definition of $\tau^{(S)}$ it is clear that $P(T^1_{\mathbf{x}} > \tau^{(S)}) \leq \frac{1}{2e}$. Moreover, by the "submultiplicativity" property (see Aldous and Fill [2], chapter 2)

$$d(s+t) \leq 2d(s)d(t), \quad s,t \geq 0, \tag{3.11}$$

we have that

$$P(T^1_{\mathbf{x}} > k\tau^{(S)}) \leq \frac{1}{2e^k}, \quad k \geq 1. \tag{3.12}$$

Thus

$$\begin{aligned}
\mathbb{E}T^1_{\mathbf{x}} &= \sum_{k=1}^{\infty} \mathbb{E}(T^1_{\mathbf{x}} \mathbf{1}((k-1)\tau^{(S)} < T^1_{\mathbf{x}} \leq k\tau^{(S)})) \\
&\leq \tau^{(S)} \sum_{k=1}^{\infty} kP((k-1)\tau^{(S)} < T^1_{\mathbf{x}} \leq k\tau^{(S)}) \\
&\leq \tau^{(S)} \sum_{k=1}^{\infty} kP((k-1)\tau^{(S)} < T^1_{\mathbf{x}}) \\
&\leq \tau^{(S)} \left(1 + \sum_{k=2}^{\infty} k\frac{1}{2e^{k-1}}\right).
\end{aligned}$$

Since the series in the right hand side converges we have

$$\mathbb{E}T^1_{\mathbf{x}} = O(f_S(d)).$$

Once the bivariate $\mathbf{S}$ process hits the diagonal

$$\mathbf{D} = \{(\mathbf{x}, \mathbf{y}) \in Q_d \times Q_d; \sum_{i=1}^{d} x_i = \sum_{i=1}^{d} y_i\}, \tag{3.13}$$

we devise one obvious coupling that forces the bivariate $\mathbf{X}$ process to stay in $\mathbf{D}$ and such that the distance defined in (3.5) between the marginal processes does not decrease. In words: we select one coordinate at random; if the marginal processes coincide in that coordinate, we allow them to evolve together; otherwise we select another coordinate in order to force two new coincidences. Formally, for each $(\mathbf{X}(t), \mathbf{Y}(t)) \in \mathbf{D}$, let I_1, I_2 and I_3 be the partition of $\{0, 1, \ldots, d\}$ such that

$$\begin{aligned}
I_1 &= \{i; X_i(t) = Y_i(t)\} \\
I_2 &= \{i; X_i(t) = 0, Y_i(t) = 1\} \\
I_3 &= \{i; X_i(t) = 1, Y_i(t) = 0\}
\end{aligned}$$

Given $(\mathbf{X}(t), \mathbf{Y}(t)) \in \mathbf{D}$, choose i u.a.r. from $\{0, 1, \ldots, d\}$.

(a) If $i \in I_1$;

1. If $X_i(t) = 1$ then make $X_i(t+1) = Y_i(t+1) = 0$ with probability $p_{s(\mathbf{X}(t))-1}$; otherwise $X_i(t+1) = Y_i(t+1) = 1$.
2. If $X_i(t) = 0$ then make $X_i(t+1) = Y_i(t+1) = 1$ with probability $p_{s(\mathbf{X}(t))}$; otherwise $X_i(t+1) = Y_i(t+1) = 0$.

(b) If $i \in I_2$;

1. Select $j \in I_3$ u.a.r.;
2. Make $X_i(t+1) = Y_j(t+1) = 1$ with probability $p_{s(\mathbf{X}(t))}$; otherwise $X_i(t+1) = Y_j(t+1) = 0$.

(c) If $i \in I_3$;

1. Select $j \in I_2$ u.a.r.;
2. Make $X_i(t+1) = Y_j(t+1) = 0$ with probability $p_{s(\mathbf{X}(t))-1}$; otherwise $X_i(t+1) = Y_j(t+1) = 1$.

Then, it is easy to check that $(\mathbf{X}(t+1), \mathbf{Y}(t+1)) \in \mathbf{D}$ and $d(\mathbf{X}(t+1), \mathbf{Y}(t+1)) \leq d(\mathbf{X}(t), \mathbf{Y}(t))$. Moreover, noticing that $|I_2| = |I_3| = d(\mathbf{X}(t), \mathbf{Y}(t))/2$, we have

$$\mathbb{P}(d(\mathbf{X}(t+1), \mathbf{Y}(t+1)) = d(\mathbf{X}(t), \mathbf{Y}(t)) - 2 \mid \mathbf{X}(t), \mathbf{Y}(t)) = \frac{d(\mathbf{X}(t), \mathbf{Y}(t))}{2d}(p_{s(\mathbf{X}(t))} + p_{s(\mathbf{X}(t))-1}) \tag{3.14}$$

$$\mathbb{P}(d(\mathbf{X}(t+1), \mathbf{Y}(t+1)) = d(\mathbf{X}(t), \mathbf{Y}(t)) \mid \mathbf{X}(t), \mathbf{Y}(t)) = 1 - \frac{d(\mathbf{X}(t), \mathbf{Y}(t))}{2d}(p_{s(\mathbf{X}(t))} + p_{s(\mathbf{X}(t))-1}). \tag{3.15}$$

In this case, it is straightforward to compute

$$\begin{aligned} m(i, s(\mathbf{X}(t))) &= i - \mathbb{E}[d(\mathbf{X}(t+1), \mathbf{Y}(t+1)) \mid d(\mathbf{X}(t), \mathbf{Y}(t)) = i, \mathbf{X}(t), \mathbf{Y}(t)] \\ &= \frac{i}{d}(p_{s(\mathbf{X}(t))} + p_{s(\mathbf{X}(t))-1}). \end{aligned} \tag{3.16}$$

Let $T^2_{\mathbf{x}}$ be the coupling time for the second coupling just described. That is, let $T^2_{\mathbf{x}} = \inf\{t > T^1_{\mathbf{x}} : d(\mathbf{X}(t), \mathbf{Y}(t)) = 0\}$. Then, as a consequence of the optional sampling theorem for martingales we have the following comparison lemma (*cf.* Aldous and Fill [2], Chapter 2).

Lemma 3.17

$$\mathbb{E}[T_{\mathbf{x}}^2 | d(\mathbf{X}(T_{\mathbf{x}}^1), \mathbf{Y}(T_{\mathbf{x}}^1)) = L, (\mathbf{X}(T_{\mathbf{x}}^1), \mathbf{Y}(T_{\mathbf{x}}^1)) \in \mathbf{D}, s(\mathbf{X}(T_{\mathbf{x}}^1)) = s] \leq \sum_{i=1}^{L} \frac{d}{i(p_s + p_{s-1})} \tag{3.18}$$

for all $s = 0, 1, \ldots, d$.

Proof. Define $(\mathbf{X}'(t), \mathbf{Y}'(t)) = (\mathbf{X}(t+T_{\mathbf{x}}^1), \mathbf{Y}(t+T_{\mathbf{x}}^1)$ for all $t \geq 0$. Define $Z(t) = d(\mathbf{X}'(t), \mathbf{Y}'(t))$ and $\mathcal{F}_t = \sigma(\mathbf{X}'(t), \mathbf{Y}'(t))$. Then, it follows from (3.16) that

$$m(i, s) = i - E[Z(1)|Z(0) = i, s(\mathbf{X}'(0)) = s]. \tag{3.19}$$

Also, for all $s \in \{1, \ldots, d-1\}$, $0 < m(1, s) \leq m(2, s) \leq \ldots \leq m(d, s)$. Fix $s \in \{1, \ldots, d-1\}$ and write

$$h(i) = \sum_{j=1}^{i} \frac{1}{m(i, s)} \tag{3.20}$$

and extend h by linear interpolation for all real $0 \leq x \leq d$. Then h is concave and for all $i \geq 1$

$$\begin{aligned} \mathbb{E}[h(Z(1)) \mid Z(0) = i, s(\mathbf{X}'(0)) = s] &\leq h(i - m(i, s)) \\ &\leq h(i) - m(i, s)h'(i) \\ &= h(i) - 1, \end{aligned}$$

where the first inequality follows from the concavity of h and h' is the first derivative of h. Now, defining $\tilde{h}$ such that

$$h(i) = 1 + \sum_{j} \mathbb{P}[h(Z(1)) \mid Z(0) = i, s(\mathbf{X}'(0)) = s]h(j) + \tilde{h}(i) \tag{3.21}$$

and

$$M(t) = t + h(Z(t)) + \sum_{u=0}^{t-1} \tilde{h}(Z(u)) \tag{3.22}$$

we have that M is an $\mathcal{F}_t$-martingale and applying the optional sampling theorem to the stopping time $T_0 = \inf\{t; Z(t) = 0\}$ we have

$$\mathbb{E}[M(T_0) \mid Z(0) = i, s(\mathbf{X}'(0)) = s] = \mathbb{E}[M(0) \mid Z(0) = i, s(\mathbf{X}'(0)) = s] = h(i). \tag{3.23}$$

Noticing that $M(T_0) \geq T_0$ and $T_0 = T^2_{\mathbf{x}}$, we obtain the desired result •

Since $s(\mathbf{X}(t))$ is distributed as π_S, we can write:

$$\mathbb{E}[T^2_{\mathbf{x}} | d(\mathbf{X}(T^1_{\mathbf{x}}), \mathbf{Y}(T^1_{\mathbf{x}})) = L, (\mathbf{X}(T^1_{\mathbf{x}}), \mathbf{Y}(T^1_{\mathbf{x}})) \in \mathbf{D}] \leq \sum_{s=0}^{d} \pi_S(s) \sum_{i=1}^{L} \frac{d}{i(p_s + p_{s-1})} := g(d). \tag{3.24}$$

Putting the pieces together, we have found a coupling time $T_{\mathbf{x}}$ for the whole process such that

$$\mathbb{E}T_{\mathbf{x}} \leq f_S(d) + g(d).$$

The task now is to find explicit bounds for $f_S(d)$ and $g(d)$ for particular workable cases.

To avoid unnecessary complications, we will assume $d = 2m$, and compute only the hitting times for the $\mathbf{S}$ process of the type $E_0 T_m$. Hitting times in birth-and-death processes assume the following closed-form (see [11] for an electrical derivation):

$$\mathbb{E}_k T_{k+1} = \frac{1}{\pi_k P(k, k+1)} \sum_{i=0}^{k} \pi_i, \qquad 0 \leq k \leq d-1,$$

and in our case this expression turns into

$$\mathbb{E}_k T_{k+1} = \frac{\sum_{i=0}^{k} \binom{d}{i}}{\binom{d-1}{k} p_k}.$$

Therefore

$$\mathbb{E}_0 T_m = \sum_{k=0}^{m-1} \frac{\sum_{i=0}^{k} \binom{2m}{i}}{\binom{2m-1}{k} p_k}. \tag{3.25}$$

(i) In case all $p_k = 1$, we have the simple random walk on the cube, and it turns out there is an even more compact expression of (3.25), namely:

$$\mathbb{E}_0 T_m = \sum_{k=0}^{m-1} \frac{\sum_{i=0}^{k} \binom{2m}{i}}{\binom{2m-1}{k}} = m \sum_{k=0}^{m-1} \frac{1}{2k+1}, \tag{3.26}$$

as was proved in [3], and the right hand side of (3.26) equals $m[H(2m) - \frac{1}{2}H(m)]$, where $H(n) = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$, allowing us to conclude immediately that in this case $\mathbb{E}_0 T_m = \mathbb{E}_0 T_{d/2} \approx \frac{d}{4} \log d + \frac{d}{4} \log 2$.

Also, we have

$$\mathbb{E}[T_{\mathbf{x}}^2|d(\mathbf{X}(T_{\mathbf{x}}^1),\mathbf{Y}(T_{\mathbf{x}}^1))=L,(\mathbf{X}(T_{\mathbf{x}}^1),\mathbf{Y}(T_{\mathbf{x}}^1))\in\mathbf{D}]\le\frac{d}{2p}\sum_{i=1}^{L}\frac{1}{i}\approx\frac{d}{2p}O(\log L). \tag{3.27}$$

Thus in this case both $f_S(d)$ and $g(d)$, and a fortiori $\mathbb{E}T_{\mathbf{x}}$ and τ, are $O(d\log d)$.

(ii) For the Aldous cube, $p_k=\frac{k}{d-1}$, and (3.25) becomes (recall this cube excludes the origin):

$$\mathbb{E}_1T_m=\sum_{k=1}^{m-1}\frac{\sum_{i=0}^{k}\binom{2m}{i}}{\binom{2m-2}{k-1}}=\sum_{k=0}^{m-2}\frac{\sum_{i=0}^{k}\binom{2m}{i}}{\binom{2m-2}{k}}+\sum_{k=0}^{m-2}\frac{\binom{2m}{k+1}}{\binom{2m-2}{k}}. \tag{3.28}$$

After some algebra, it can be shown that the second summand in (3.28) equals $(2m-1)[H(2m-1)-\dfrac{1}{m}]$, and the first summand can be bounded by twice the expression in (3.26), on account of the fact that $\binom{2m-1}{k}\le 2\binom{2m-2}{k}$, for $0\le k\le m-1$. Therefore, we can write

$$\mathbb{E}_1T_{d/2}\le\frac{3}{2}d\log d+\ \text{smaller terms},$$

thus improving by a factor of $\frac{1}{2}$ the computation of the same hitting time in [5].

Also, we have

$$\mathbb{E}[T_{\mathbf{x}}^2|d(\mathbf{X}(T_{\mathbf{x}}^1),\mathbf{Y}(T_{\mathbf{x}}^1))=L,(\mathbf{X}(T_{\mathbf{x}}^1),\mathbf{Y}(T_{\mathbf{x}}^1))\in\mathbf{D},s(\mathbf{X}(T_{\mathbf{x}}^1))=s]\le\sum_{i=1}^{L}\frac{d(d-1)}{i(2s-1)} \tag{3.29}$$

Thus, in this case

$$\begin{aligned}
&\mathbb{E}[T_{\mathbf{x}}^2|d(\mathbf{X}(T_{\mathbf{x}}^1),\mathbf{Y}(T_{\mathbf{x}}^1))=d,(\mathbf{X}(T_{\mathbf{x}}^1),\mathbf{Y}(T_{\mathbf{x}}^1))\in\mathbf{D}]\\
&\quad\le\ \sum_{s=1}^{d}\pi_S(s)\sum_{i=1}^{d}\frac{d(d-1)}{i(2s-1)}\\
&\quad\le\ \Phi(-\sqrt{d}/3)\sum_{i=1}^{d}\frac{d(d-1)}{i}+(1-\Phi(-\sqrt{d}/3))\sum_{i=1}^{d}\frac{d(d-1)}{i(2d/3-1)}\\
&\quad\approx\ e^{-d/9}d(d-1)\log d+(1-e^{-d/9})d\log d.
\end{aligned} \tag{3.30}$$

And so $\tau=O(d\log d)$ also in this case.

(iii) Slower walks. Consider the case when the probability p_k grows exponentially in k, more specifically

$$p_k = \Big(\frac{k+1}{n+1}\Big)^{\alpha} \tag{3.31}$$

with $\alpha > 1$. In this case, it seems that (3.25) is useless to get a closed expression for $\mathbb{E}_0 T_{d/2}$. However, Graham and Chung [5] provide the following bound

$$\mathbb{E}_i T_{d/2} \le c_0(\alpha) d^{\alpha}, \qquad \text{for all } d \ge d_0(\alpha), 0 \le i \le d \tag{3.32}$$

where $c_0(\alpha)$ and $d_0(\alpha)$ are constants depending only on α. Moreover, (3.24) becomes

$$\begin{aligned} g(d) &= \sum_{s=0}^{d} \pi_S(s) \sum_{i=1}^{d} \frac{d(d+1)^{\alpha}}{i((s+1)^{\alpha} + s^{\alpha})} \\ &= d(d+1)^{\alpha} \sum_{s=0}^{d} \frac{\pi_S(s)}{(s+1)^{\alpha} + s^{\alpha}} \sum_{i=1}^{d} \frac{1}{i} \\ &\approx d(d+1)^{\alpha} \log d \sum_{s=0}^{d} \frac{\pi_S(s)}{(s+1)^{\alpha} + s^{\alpha}} \\ &\le d(d+1)^{\alpha} \log d \sum_{s=0}^{d} \frac{\pi_S(s)}{(s+1)^{\alpha}} \\ &= d(d+1)^{\alpha} \log d \mathbb{E}\Big[\frac{1}{(1+X)^{\alpha}}\Big], \end{aligned} \tag{3.33}$$

where X is a Binomial $(d, \frac{1}{2})$ random variable. Jensen's inequality and the same argument that lead to (3.30) show that $\mathbb{E}[(1+X)^{-\alpha}] \sim O(d^{-\alpha})$ and (3.33) can be bound by $O(d \log d)$. This fact together with (3.32) allows us to conclude that $\tau = O(d^{\alpha})$ in this case, thus improving on the rate of the mixing time provided in [5] by a factor of $\log d$.

Acknowledgments. This paper was initiated when both authors were visiting scholars at the Statistics Department of the University of California at Berkeley. The authors wish to thank their hosts, especially Dr. David Aldous with whom they shared many fruitful discussions. This work was partially supported by FAPESP 99/00260-3.

References

[1] Aldous, D.J. (1982) Some inqualities for reversible Markov chains, *Journal of the London Mathematical Society*, **2**, 25:564–576.

[2] Aldous, D. J. and Fill, J. (2000) Reversible Markov chains and random walks on graphs, book draft.

[3] Blom, G. (1989) Mean transition times for the Ehrenfest urn, *Advances in Applied Probability*, **21**, 479-480.

[4] Chandra, A. K., P. Raghavan, W. L. Ruzzo, R. Smolensky and P. Tiwari (1989), The electrical resistance of a graph captures its commute and cover times, in *Proceedings of the Twenty First Annual ACM Symposium on Theory of Computing*, Seattle, Washington, pp. 574-586.

[5] Chung, F. R. K. and Graham, R. L. (1997) Stratified random walks on the n-cube, *Random Structures and Algorithms*, **11**, 199-222.

[6] Feige, U. (1995)) A tight lower bound on the cover time for random walks on graphs, *Random Structures and Algorithms*, **6**, 433-438.

[7] Griffeath, D. (1975) A maximal coupling for Markov chains, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **31**, 95-106.

[8] Lovász, L. and Winkler, P. (1998) Mixing times, in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, **41**, American Mathematical Society.

[9] Matthews, P. C. (1988) Covering problems for Markov chains, *Annals of Probability*, **16**, 1215-1228.

[10] Palacios, J. L. (1994) Another look at the Ehrenfest urn via electric networks, *Advances in Applied Probability*, **26**, 820-824.

[11] Palacios, J. L. and Tetali, P. (1996) A note on expected hitting times for birth and death chains, *Statistics & Probability Letters*, **30**, 119-125.